



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Meteorological conditions and incidence of Legionnaires' disease in Glasgow, Scotland: application of statistical modelling

Citation for published version:

Bhopal, R, Dunn, CE, Rowlingson, B & Diggle, P 2013, 'Meteorological conditions and incidence of Legionnaires' disease in Glasgow, Scotland: application of statistical modelling', *Epidemiology & Infection*, vol. 141, no. 4, pp. 687-696. <https://doi.org/10.1017/S095026881200101X>

Digital Object Identifier (DOI):

[10.1017/S095026881200101X](https://doi.org/10.1017/S095026881200101X)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Epidemiology & Infection

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Meteorological conditions and incidence of Legionnaires' disease in Glasgow, Scotland: application of statistical modelling

C. E. DUNN¹*, B. ROWLINGSON², R. S. BHOPAL³ AND P. DIGGLE⁴

¹ Department of Geography, Durham University, Durham, UK

² Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

³ Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, UK

⁴ School of Health and Medicine, Lancaster University, Lancaster, UK

Received 4 November 2011; Final revision 18 April 2012; Accepted 2 May 2012

SUMMARY

This study investigated the relationships between Legionnaires' disease (LD) incidence and weather in Glasgow, UK, by using advanced statistical methods. Using daily meteorological data and 78 LD cases with known exact date of onset, we fitted a series of Poisson log-linear regression models with explanatory variables for air temperature, relative humidity, wind speed and year, and sine-cosine terms for within-year seasonal variation. Our initial model showed an association between LD incidence and 2-day lagged humidity (positive, $P=0.0236$) and wind speed (negative, $P=0.033$). However, after adjusting for year-by-year and seasonal variation in cases there were no significant associations with weather. We also used normal linear models to assess the importance of short-term, unseasonable weather values. The most significant association was between LD incidence and air temperature residual lagged by 1 day prior to onset ($P=0.0014$). The contextual role of unseasonably high air temperatures is worthy of further investigation. Our methods and results have further advanced understanding of the role which weather plays in risk of LD infection.

Key words: Legionnaires' disease, statistics.

INTRODUCTION

The principal sources of infection for Legionnaires' disease (LD), an environmentally acquired pneumonia contracted through inhalation of aerosols contaminated with *Legionella* bacteria, have been identified as human-made, complex reservoirs of warm recirculated water such as cooling towers and complex hot-water systems [1–3], evaporative

condensers [4, 5], whirlpool spas [6, 7] and showers [8]. The specific environmental conditions under which contaminated aerosol is delivered at an infective dose, sometimes several hundred, or even thousand, metres from the source [4, 9, 10], however, remain poorly understood. Community-acquired LD, both in sporadic (i.e. cases not part of a recognized cluster) and outbreak form, shows marked spatial and temporal variations in incidence. Our previous work in Glasgow, Scotland, for example, has shown a multiplicity of clusters in time and space, and an association between disease incidence and distance of home residence from a cooling tower [9, 11]. Until

* Author for correspondence: Dr C. E. Dunn, Department of Geography, Durham University, South Road, Durham, DH1 3LE, UK.
(Email: C.E.Dunn@durham.ac.uk)

recently little attention has been given to the specific meteorological conditions which are optimal for transport and delivery of an infective dose of *Legionella*-contaminated aerosol. Earlier analysis of data for Glasgow shows a higher incidence of LD in autumn and early winter [12] which contrasts with more recent findings for the northeastern USA [13], England and Wales [14], and The Netherlands [15, 16] where cases have been reported to occur predominantly in summer. Karagiannis *et al.* [16] found that warm (but not the hottest), humid weather conditions in summer in The Netherlands coincided with the highest LD incidence. Hicks *et al.* [17] recently reported that cases of legionellosis in the USA for the decade 2000–2009 occurred mostly in summer and early autumn (June–October) and, similarly, a study in Toronto, Canada showed legionellosis cases to peak during late summer or early autumn [18]. Cooling tower-associated outbreaks in Australia have been shown to be most frequent in April (autumn in the southern hemisphere) [19].

More specifically Fisman *et al.* [13] demonstrated that for the Philadelphia metropolitan region of the USA incidence of legionellosis was positively associated with relative humidity and precipitation, and negatively associated with average wind speed in the period 6–10 days before onset. García-Fulgueiras *et al.* [20] noted how low wind speeds (average 9 kph) in Murcia, Spain engendered dispersal of aerosols while other work has suggested a relationship between *Legionella* survival/LD incidence and increased humidity [4, 14, 21] and between legionellosis incidence and increased rainfall [22]. In contrast, and by measuring concentrations of *Legionella* species in air samples around aeration ponds at a biological treatment plant in Norway, Blatny *et al.* [23] found the highest concentrations during cloudy weather and the lowest during rain. Sala Ferré *et al.* [24] suggest that relatively high temperatures (mean 14.4 °C), high humidity (mean 83%) and low mean wind speed (3.6 kph), as well as flat terrain for an area north of Barcelona, Spain could help to explain airborne dispersal of contaminated aerosol from an industrial cooling tower and hence a community outbreak in October 2005. Ng *et al.* [18] found evidence for an increase in odds of legionellosis occurring with increasing humidity in Toronto, Canada but they identified local hydrological changes, particularly decreases in river flow, to be the strongest contributors to disease risk.

AIMS AND DATASETS

The aim of this paper is to show how statistical modelling can be used to investigate the potential short-term associations between LD occurrence in Glasgow, Scotland and local meteorological conditions. Based on findings cited in the literature above, the epidemiology of LD [4, 11, 25], laboratory data [21] and the specific seasonal patterns found in Glasgow, we focus on three meteorological variables: air temperature, wind speed and relative humidity. All model fitting was performed using the software R, version 2.12 [26] and freely available add-on packages (the R code for the analyses is available from the authors; however, the case data are not freely distributable).

We used a subset of the case data reported on previously [11] for patients diagnosed with community-acquired, non-travel, non-outbreak LD in Glasgow. As part of this earlier work a list of cases was compiled and validated using a range of external information and a case definition conforming to widely agreed principles [27]. The case data are therefore drawn from a historical, *ad hoc* and, for Scotland, unusual study which collected data of sufficient detail and resolution to allow the application of the advanced statistical methods discussed here.

During the period for which case data were available, 1979–1986, there were 120 cases with a Glasgow postcode. Age, sex and grid reference of place of residence were also available for these cases. Of these 120 cases, 78 had a known exact date of onset and these formed our case dataset.

Daily weather data were made available for the period of interest for Glasgow airport, located about 7 miles from the city centre. The main variables of interest comprised daily relative humidity (per cent), daily mean air temperature (tenths of 1 °C, and derived as the mean of the daily maximum and the nightly minimum) and daily mean wind speed (tenths of a knot, 0–24 h GMT).

STATISTICAL MODELLING AND RESULTS

Missing values

Of the 42 LD cases which were excluded from our analysis, month and year of onset were available for 41 and year only for the remaining case. The age range for these 42 cases was 14–84 years (mean 58.0 years) and there were 28 males (66.67%) and

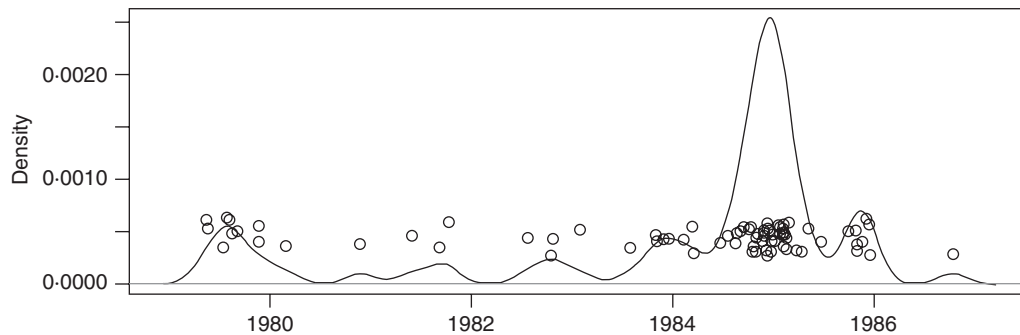


Fig. 1. Illustrative kernel-smoothing estimate of case density with superimposed jittered cases, using a kernel bandwidth = 52 days.

14 females (33.33%). This compares with an age range of 25–81 years (mean 54.7 years) and a male:female ratio of 55:23 (70.51/29.49%) for the 78 cases considered here. In terms of temporal proportions, cases without exact date of onset ($n=42$) occurred most frequently in the years 1983 (26.2%) and 1984 (21.4%) and, for cases with only month and year available ($n=41$), in the months of September (17.07%) and November (26.83%); for the 78 cases used here corresponding years and months are 1984 (38.5%) and 1985 (30.8%), and October (16.67%), November (15.38%) and December (17.95%).

Temporal patterns

For the 8-year period of interest, 1 January 1979 to 31 December 1986 ($n=2922$ days) 75 days were characterized by the occurrence of at least one LD case: 2847 days had no cases, 72 days had one case each and 3 days had two cases each. Using kernel smoothing [28] we examined the detailed temporal distribution of the 78 LD cases; the resulting kernel-smoothed density curve is shown in Figure 1. For presentation purposes the y-coordinates of the case points are shown in this figure as random offsets ('jitter') to avoid overlap. The distribution indicates temporal variation and a large increase in cases during the winter of 1984/1985.

Figure 2 shows the temporal variation in the weather variables with the case data displayed as number of LD cases per month. Strong seasonal patterns are shown for relative humidity and air temperature, roughly anti-correlated, although there is no obvious visual pattern for the two wind variables.

Statistical modelling

We examined the distribution of the 78 LD cases by meteorological conditions in the incubation period

(2–10 days before clinical onset of disease, median 6 days) by using a number of different statistical models. We first calculated a moving average (MA) of the weather variables over a 9-day window and fitted a series of Poisson log-linear regression models [29] to the number of cases on each day (including the days with 0 cases). The results of fitting the first of these models to the number of cases on each day for temperature (Temp_{MA}), relative humidity (RH_{MA}) and wind speed (SWind_{MA}), each lagged by 2 days are shown in Table 1. This shows the association of cases with the relative humidity and wind speed variables to be significantly non-zero. There is a positive association with increasing humidity and a negative association with increasing wind speed.

In order to consider the effect of the large number of cases in the winter of 1984/1985, and also the seasonal variation within each year, we extended the first model to examine periodicity in the data. This also allowed for any changes over time in diagnostic and reporting procedures for LD cases. In this model we included a categorical variable defined as the 12-month period centred around 1 January of each year, while the overall yearly rise and fall was modelled with a sine wave of period 1 year. This is parameterized with sine and cosine terms giving two parameters which can be converted to amplitude and phase. The results of this analysis (not shown) demonstrate that only the period centred on January 1985 is significant, confirming the visual impression from Figure 1. Our 'baseline' model therefore comprises an intercept and a variable that is 1 for the period around January 1985 and 0 at all other times (a TRUE/FALSE term for cases centred on the period 1 January 1985). The model results are highly significant, reflecting the increase in number of cases in the 1985 period.

When we extend the baseline model to incorporate a seasonal effect by using sine and cosine terms the

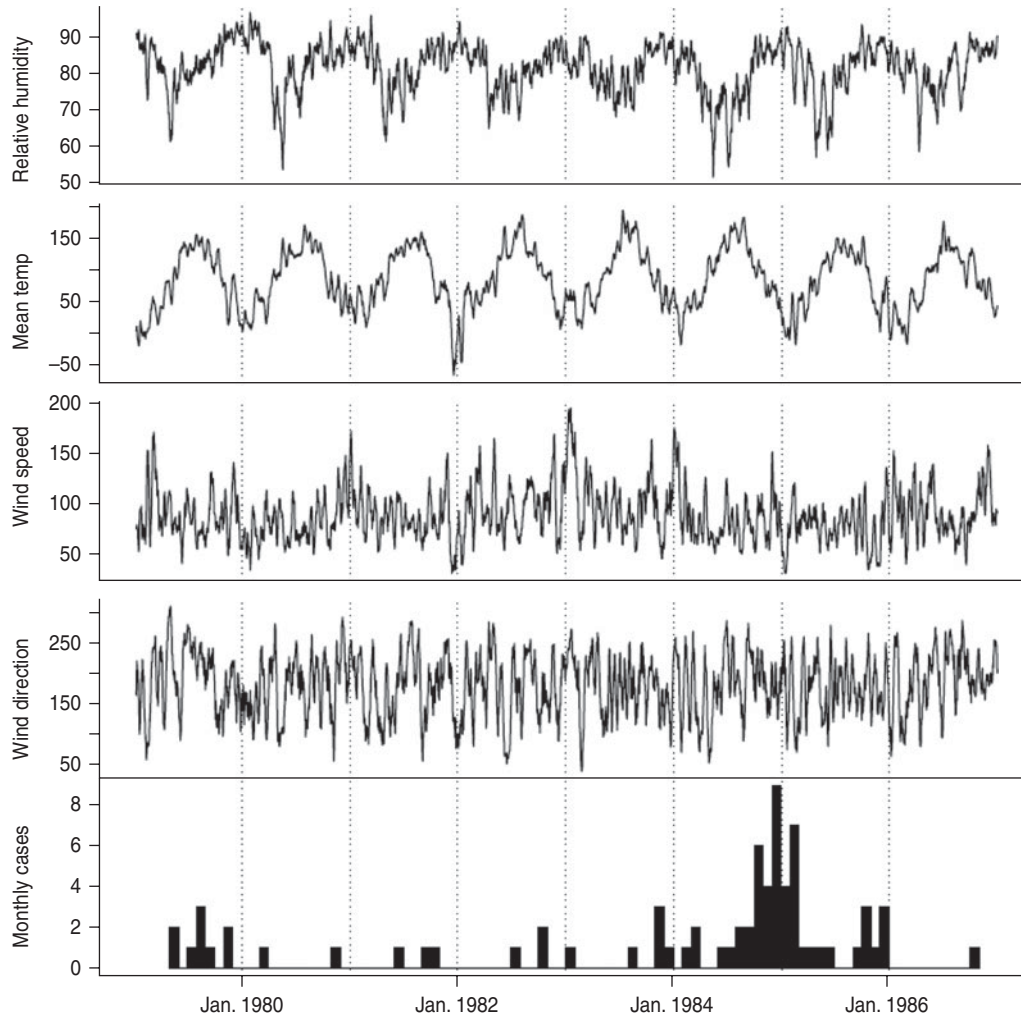


Fig. 2. Temporal variations in daily relative humidity (per cent), daily mean air temperature (tenths of 1 °C), wind speed (tenths of a knot) and wind direction (degrees) with number of Legionnaires' disease cases per month.

parameters are again highly significant as would be expected from the pattern of cases. By adding into the baseline model the three weather variables using the 2-day lagged MA, relative humidity is shown to be significant ($P=0.0005$). To complete this phase of the modelling, we extended the baseline model to include both seasonality and the weather variables (Table 2). This showed no significant relationships with weather and a weaker relationship with the sine-cosine term.

Inclusion of either weather or sine-cosine terms provides a significantly better fit than does the baseline model, confirming within-year seasonality, while inclusion of both weather and sine-cosine variables gives a significantly better fit than any of the preceding models. However, weather and sine-cosine effects are partially confounded and we need to examine whether short-term fluctuations in weather about

their longer-term seasonal patterns show association with LD incidence after adjusting for long-term seasonality in incidence by using the sine-cosine terms. We therefore fitted a simple sine-cosine model to the weather variables so that we could use the residuals as measures of these short-term fluctuations. Note that the responses in these models are the continuous weather variables, so we used normal linear models and not Poisson regression as we do when number of cases is the response. We also used the raw weather values instead of the smoothed moving averages. The residuals from these models demonstrate considerable autocorrelations which correspond to the time-scales over which weather differs consistently from the 'norm' for that time of year. In effect these are weather 'anomalies' which may include, for example, short 'cold spells' or a few windy days. Table 3 shows the results of a model

Table 1. *Poisson model for Legionnaires' diseases cases with air temperature, relative humidity and wind speed, with incident rate ratio (IRR) and 95% confidence intervals (CI)*

	Estimate	S.E.	Pr(> z)	IRR	(95% CI)	P
(Intercept)	-6.4557	1.8270	0.0004	0.0016	(0.0000-0.0607)	***
Lag (Temp _{MA} , 2)	-0.0004	0.0025	0.8864	0.9996	(0.9947-1.0047)	
Lag (RH _{MA} , 2)	0.0452	0.0200	0.0236	1.0463	(1.0053-1.0889)	*
Lag (SWind _{MA} , 2)	-0.0101	0.0047	0.0330	0.9899	(0.9806-0.9994)	*

S.E., Standard error; MA, moving average.

Deviances: 573.0201 [2911 degrees of freedom (D.F.)] to 560.1568 (2908 D.F.).

Change: 12.8633 on 3 D.F., $P=0.0049$.

Table 2. *Poisson model for cases with seasonal effect and weather variables, showing incident rate ratio (IRR) and 95% confidence intervals (CI)*

	Estimate	S.E.	Pr(> z)	IRR	(95% CI)	P
(Intercept)	-5.2109	2.3220	0.0248	0.0055	(0.0001-0.5673)	*
Year 1985	1.9989	0.2463	0.0000	7.3810	(4.5100-12.0797)	***
Sin (time)	-0.2539	0.2639	0.3359	0.7757	(0.4577-1.3149)	
Cos (time)	1.1048	0.4837	0.0224	3.0187	(1.1473-7.9426)	*
Lag (RH _{MA} , 2)	0.0107	0.0268	0.6892	1.0108	(0.9580-1.0665)	
Lag (SWind _{MA} , 2)	-0.0092	0.0056	0.1015	0.9908	(0.9798-1.0020)	
Lag (Temp _{MA} , 2)	0.0083	0.0065	0.2005	1.0084	(0.9953-1.0215)	

S.E., Standard error; MA, moving average.

Deviances: 573.0201 [2911 degrees of freedom (D.F.)] to 467.0563 (2905 D.F.).

Change: 105.9638 on 6 D.F., $P<0.001$.

which incorporates the 1985 effect, a sine-cosine term and the residuals from the weather variable models.

Finally, in order to incorporate time-lagged weather effects we fitted the model with a range of time lags. In this model the sine-cosine and 1985 effect parameters do not change significantly. We plotted the parameters of the three weather variables with 2 standard error intervals against a range of 12 lags in order to demonstrate significant non-zero values in the incubation period (Fig. 3). Positive lag values (1-9) represent days immediately prior to date of disease onset and, for illustrative purposes, negative lags (-1 and -2) correspond to lead times where the value of the predictive variable is taken for the days immediately *after* date of onset. The most significant correlation is that for temperature at a lag of 1 day prior to onset which gives an increase in incidence with increasing positive residual (Fig. 3). There is also weaker significance at 3 days prior to onset which persists at 0 days and at small lead times. Relative humidity and wind speed are much less significant and are therefore removed from the analysis to produce a final model. As shown in Table 4 there is a significant association between LD incidence and air

temperature residual lagged by 1 day prior to onset ($P=0.0014$). The residuals in this final model are uncorrelated and the residual deviance is much smaller than the residual degrees of freedom, indicating an absence of over-dispersion. None of the other Poisson models shows over-dispersion.

DISCUSSION

The statistical approach developed in this paper has further advanced understanding of the role which weather plays in risk of LD. Our initial model showed that relative humidity was positively associated and wind speed was negatively associated with LD incidence in Glasgow, Scotland. This finding is consistent with work elsewhere [4, 13-14, 20-21, 24] and lends support to the suggestion that high levels of humidity favour survival of *Legionella* bacteria [4, 21]. However, these associations were not significant once we adjusted for season and year.

We also explored the potential influence of short-term, unseasonable weather conditions or 'anomalies' and our findings show a small but significant association between LD incidence and mean air

Table 3. *Poisson model for cases with 1985 effect, sine-cosine term and residuals from weather variable models, showing incident rate ratio (IRR) and 95% confidence intervals (CI)*

	Estimate	S.E.	Pr(> z)	IRR	(95% CI)	P
(Intercept)	-4.4715	0.1840	0.0000	0.0114	(0.0079–0.0165)	***
Year 1985	2.0022	0.2327	0.0000	7.4057	(4.6503–11.7937)	***
Sin (time)	-0.5278	0.1751	0.0026	0.5899	(0.4156–0.8373)	**
Cos (time)	0.6823	0.1762	0.0001	1.9784	(1.3907–2.8145)	***
RH_residual	-0.0026	0.0128	0.8383	0.9974	(0.9721–1.0233)	
Temp_residual	0.0093	0.0041	0.0246	1.0093	(1.0010–1.0177)	*
Swind_residual	-0.0059	0.0031	0.0537	0.9941	(0.9881–1.0002)	

S.E., Standard error.

Deviances: 573.3412 [2917 degrees of freedom (D.F.)] to 464.8344 (2911 D.F.).

Change: 108.5067 on 6 D.F., $P < 0.001$.

temperature residual on the day preceding date of onset. Given the incubation period of 2–10 days prior to onset this result needs to be interpreted with careful consideration and we cannot say whether this association is causal. The association with mean air temperature residual at 3 days prior to, and immediately following, onset is almost certainly a side-effect of autocorrelation in the residual temperature values. When two time-series are autocorrelated, a direct relationship between the two at a particular lag typically generates indirect associations at other lags; hence the overall pattern of the cross-correlations is more important than their individual significance. In light of this the contextual role of unseasonably high air temperatures is therefore worthy of further investigation. Previous findings on the importance of air temperature indicate a potential role for higher seasonal temperatures although evidence is lacking for an association during the incubation period. Using mean monthly temperatures for states in the Mid-Atlantic region of the USA, Hicks *et al.* [22] found statistically significant correlations between monthly case counts of legionellosis and temperature in the preceding month. Using negative binomial regression and controlling for state and year they reported a 2.8% increase in risk for legionellosis for a 1 °C increase in temperature. In the study by Karagiannis *et al.* [16] mean weekly temperature was found to be associated with LD incidence in The Netherlands, with higher temperatures in the 2 weeks prior to exposure further contributing to higher incidence. For England and Wales Ricketts *et al.* [14] found a correlation between disease incidence and temperature 10–14 weeks prior to onset although this was dependent on both year and season (quarter); no consistently significant effect of temperature was shown for

the incubation period. Fisman *et al.* [13] similarly found no relationship between temperature during the incubation period and legionellosis incidence.

Our understanding of the complex relationships between ambient weather and risk of LD is complicated by the use of data of different spatial and temporal resolutions and quality in different settings, and in this context it is important to consider each study's limitations. First, in the present research we made the assumption that weather monitored for Glasgow airport pertains to the whole of the Glasgow postcode area. Although the airport weather station was the nearest available station monitoring daily data, its use may not account for smaller scale spatial variations within the urban environment. Additionally, and as indicated by Ricketts *et al.* [14], even more highly localized variations in aerosol concentration can occur for sources of *Legionella* bacteria such as spa pools. It is therefore difficult to comprehensively account for each and every variation in the meteorological conditions of a complex environment such as a large city. Further work is warranted to explore spatial dispersal mechanisms through techniques such as air quality dispersion modelling, the results of which can be interfaced to Geographical Information Systems (GIS) software [10, 30, 31]. However, the nature of aerosol emissions from sources such as cooling towers and condensers may present greater complexities than, say, a single factory chimney whose emissions can be more readily modelled. Blatny *et al.* [23] propose the use of computational fluid dynamics to estimate aerosol paths.

Second, suitability of the individual daily meteorological variables needs to be considered. Temporal dispersion of aerosols through the ambient atmosphere is complex, cooling tower mists being carried in

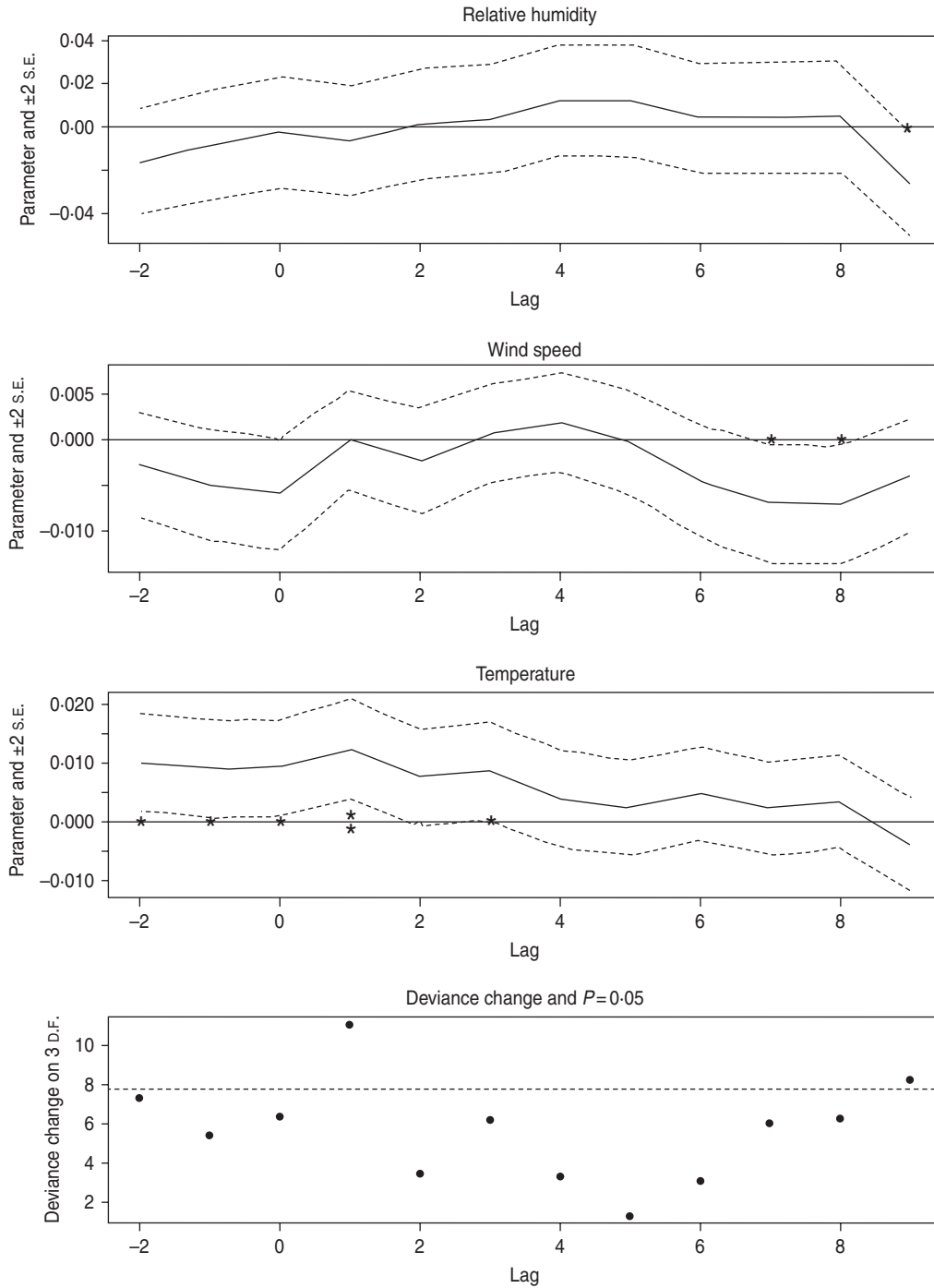


Fig. 3. Weather variable parameters plotted with 2 standard errors (S.E.) against time lags of -2 to 9 days (* $P=0.05$, ** $P=0.01$).

different directions at different times of the day [23, 32]. It should therefore be considered that variations in weather patterns are possibly not adequately characterized by summary daily measures such as daily means which we used in our analyses of wind speed and air temperature. Similarly, relative humidity is measured as a single value recorded at

09:00 hours GMT which may hide diurnal variations caused, for example, by changes in temperature or by advection of air masses. These summary measures, therefore, may not be fully representative of the ambient conditions which determine aerosol dispersion.

Third, it is important to consider potential errors in diagnosis and reporting of LD cases. Our analysis

Table 4. *Poisson model for cases with 1985 effect, sine-cosine term and residuals for temperature lagged by one day, showing Incident Rate Ratio (IRR) and 95 % confidence intervals (CI)*

	Estimate	S.E.	Pr(> z)	IRR	95 % CI	P
(Intercept)	-4.5027	0.1858	0.0000	0.0111	(0.0076-0.0161)	***
Year 1985	2.0702	0.2273	0.0000	7.9261	(5.0307-12.4878)	***
Sin (time)	-0.5086	0.1749	0.0036	0.6013	(0.4238-0.8532)	**
Cos (time)	0.6592	0.1764	0.0002	1.9332	(1.3586-2.7508)	***
Lag-1 Temp_residual	0.0124	0.0039	0.0014	1.0125	(1.0047-1.0204)	**

Deviances: 573.5015 (2920 D.F.) to 460.3270 (2916 D.F.).

Change: 113.1745 on 4 D.F., $P < 0.001$.

is based on a relatively small number of cases and pertains to a historical time period. However, although dated, the dataset has been validated as part of earlier work and meets the objective of the present study to use a number of different statistical models to examine the potential relationship with meteorological conditions pertaining to the time of diagnosis. There is, however, a need for analysis of more contemporary case data for Scotland and also for our methods to be applied to datasets for other settings, and to this end we welcome collaborative opportunities.

In addition to the 78 cases which form the subject of analysis here, an additional 42 cases were excluded because the exact date of onset was missing. This level of completion (65%) is not unexpected given the historical nature of the dataset. For more current data Fisman *et al.* [13], for instance, report that 85% of legionellosis cases had a documented date of symptoms onset. The case dataset shows a significantly large number of LD cases in the winter of 1984/1985. There was an outbreak of LD in Glasgow in summer 1984 [33] and it is possible that heightened awareness and increased vigilance may have resulted in additional diagnosis and reporting of cases in subsequent months. An alternative explanation may relate to the possibility of a more highly pathogenic strain of the bacterium being present at that time [14, 34]. The explanation for this increase notwithstanding, our analysis has attempted to account for the excessively large number of cases which occurred at this time by including a factor for this period.

Finally, it is important to consider the spatial boundary of the case data. We have limited our dataset to cases which were identified by a Glasgow area (G) home address postcode. Of other cases in Scotland the postcode area PA (Paisley) is located close to Glasgow city (to the north and west) although the number of cases in this postcode area which were

categorized as community-acquired, non-travel, non-outbreak and with an exact date of onset is very small ($n = 3$).

Recent years have witnessed a growth in interest in investigating the meteorological conditions which may be associated with raised incidences of LD. Identification of such conditions is important since it not only facilitates our understanding of the environments in which *Legionella* bacteria survive and are dispersed, but may also aid disease detection and prediction. Findings from this research avenue are likely to become increasingly pertinent to public health policy and practice given changes in global climate [14, 16]. Our approach in this paper has highlighted how cross-disciplinary collaboration provides a valuable platform for enhancing our understanding of the specific meteorological conditions which may be most pertinent under such changing climates.

Despite, or perhaps because of, the recent flurry of research activity around the relationship between ambient weather conditions and the onset of LD a number of different findings have emerged as well as some similarities. Our analysis of daily meteorological data for Glasgow, Scotland has indicated a potentially complex relationship between ambient weather conditions and LD incidence. There is evidence of temporal variation over an 8-year period with a large peak in cases in one particular winter. Although our initial, simple models indicated a significant association between disease incidence and both relative humidity and wind speed conditions during the preceding days of incubation, further developments to account for seasonal effects show no significant relationships with weather variables. After adjusting for long-term seasonality in disease incidence, short-term changes in weather were examined for potential associations with disease incidence by using the residuals from the weather variable models; this

demonstrates a significant association with the mean air temperature residual, although most notably outside of the incubation period, and not for relative humidity or wind speed. Fisman *et al.*'s [13, p. 2072] statement that 'the seasonality of non-vectorborne infectious diseases remains poorly understood' is still relevant, therefore, and there is need for additional detailed analysis of the role of the three weather variables considered here. By adding to a sparse but important area of research in airborne infections our methods and findings form a strong foundation for such further work.

ACKNOWLEDGEMENTS

The Legionnaires' disease dataset was collected with financial support from the former Greater Glasgow Health Board and the former Pneumonia Research Trust. We thank the late Dr R. J. Fallon for his invaluable role in establishing a register of cases. We thank the patients, general practitioners, hospital consultants and medical records officers for help with compiling the database. We thank the Glasgow Weather Centre for meteorological data.

DECLARATION OF INTEREST

None.

REFERENCES

1. Bhopal RS. Geographical variation of Legionnaires' disease: a critique and guide to future research. *International Journal of Epidemiology* 1993; **22**: 1127–1136.
2. Bhopal RS, Wagstaff R. Prospects for the elimination of Legionnaires' disease. *Journal of Infection* 1993; **26**: 239–243.
3. Dondero Jr. TJ, *et al.* An outbreak of Legionnaires' disease associated with a contaminated air-conditioning cooling tower. *New England Journal of Medicine* 1980; **302**: 365–370.
4. Addiss DG, *et al.* Community-acquired Legionnaires' disease associated with a cooling tower: evidence for longer-distance transport of *Legionella pneumophila*. *American Journal of Epidemiology* 1989; **130**: 557–568.
5. Breiman RF, *et al.* Role of air sampling in investigation of an outbreak of Legionnaires' disease associated with exposure to aerosols from an evaporative condenser. *Journal of Infectious Diseases* 1990; **161**: 1257–1261.
6. Benkel DH, *et al.* Outbreak of Legionnaires' disease associated with a display whirlpool spa. *International Journal of Epidemiology* 2000; **29**: 1092–1098.
7. Jernigan DB, *et al.* Outbreak of Legionnaires' disease among cruise ship passengers exposed to a contaminated whirlpool spa. *Lancet* 1996; **347**: 494–499.
8. Breiman RF, *et al.* Association of shower use with Legionnaires' disease: possible role of amoebae. *Journal of American Medical Association* 1990; **263**: 2924–2926.
9. Dunn CE, *et al.* Advancing insights into methods for studying environment–health relationships: a multi-disciplinary approach to understanding Legionnaires' disease. *Health & Place* 2007; **13**: 677–690.
10. Nguyen TMN, *et al.* A community-wide outbreak of legionnaires disease linked to industrial cooling towers – how far can contaminated aerosols spread? *Journal of Infectious Diseases* 2006; **193**: 102–111.
11. Bhopal RS, Diggle P, Rowlingson B. Pinpointing clusters of apparently sporadic cases of legionnaires' disease. *British Medical Journal* 1992; **304**: 1022–1027.
12. Bhopal RS, Fallon RJ. Seasonal variation of Legionnaires' disease in Scotland. *Journal of Infection* 1991; **22**: 153–160.
13. Fisman DN, *et al.* It's not the heat, it's the humidity: wet weather increases legionellosis risk in the greater Philadelphia metropolitan area. *Journal of Infectious Diseases* 2005; **192**: 2066–2073.
14. Ricketts KD, *et al.* Weather patterns and Legionnaires' disease: a meteorological study. *Epidemiology and Infection* 2009; **137**: 1003–1012.
15. Joseph C, van der Sande M. Unexplained summer increase in non-travel related legionellosis in the UK and Netherlands. *Eurosurveillance* 2006; **11**.
16. Karagiannis I, Brandsema P, van der Sande M. Warm, wet weather associated with increased Legionnaires' disease incidence in The Netherlands. *Epidemiology and Infection* 2009; **137**: 181–187.
17. Hicks LA, *et al.* Legionellosis – United States, 2000–2009. *Journal of the American Medical Association* 2011; **306**: 1645–1647.
18. Ng V, *et al.* Going with the flow: legionellosis risk in Toronto, Canada is strongly associated with local watershed hydrology. *EcoHealth* 2008; **5**: 482–490.
19. Bentham RH, Broadbent CR. A model for autumn outbreaks of Legionnaires' disease associated with cooling towers, linked to system operation and size. *Epidemiology and Infection* 1993; **111**: 287–295.
20. García-Fulgueiras A, *et al.* Legionnaires' disease outbreak in Murcia, Spain. *Emerging Infectious Diseases* 2003; **9**: 915–921.
21. Berendt RF. Survival of *Legionella pneumophila* in aerosols: effect of relative humidity. *Journal of Infectious Diseases* 1980; **141**: 689.
22. Hicks LA, *et al.* Increased rainfall is associated with increased risk for legionellosis. *Epidemiology and Infection* 2007; **135**: 811–817.
23. Blatny JM, *et al.* Tracking airborne *Legionella* and *Legionella pneumophila* at a biological treatment plant. *Environmental Science & Technology* 2008; **42**: 7360–7367.
24. Sala Ferré MR, *et al.* A community outbreak of Legionnaires' disease associated with a cooling tower

- in Vic and Gurb, Catalonia (Spain) in 2005. *European Journal of Clinical Microbiology & Infectious Diseases* 2009; **28**: 153–159.
25. **O'Brien SJ, Bhopal RS.** Legionnaires' disease: the infective dose paradox. *Lancet* 1993; **342**: 5–6.
 26. **R Development Core Team.** *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2011.
 27. **Bhopal RS, et al.** Proximity of the home to a cooling tower and risk of non-outbreak legionnaires' disease. *British Medical Journal* 1991; **302**: 378–383.
 28. **Silverman BW.** *Density Estimation for Statistics and Data Analysis*. Boca Raton, Florida: Chapman and Hall/CRC Press, 1992, pp. 176.
 29. **McCullagh P, Nelder JA.** *Generalized Linear Models*. Boca Raton, Florida: Chapman and Hall/CRC Press, 1989, pp. 532.
 30. **Dunn CE, Kingham S.** Establishing links between air quality and health: searching for the impossible? *Social Science and Medicine* 1996; **42**: 831–841.
 31. **Dunn CE, et al.** Analysing spatially referenced public health data: a comparison of three methodological approaches. *Health & Place* 2001; **7**: 1–12.
 32. **Brown CM, et al.** A community outbreak of Legionnaires' disease linked to hospital cooling towers: an epidemiological method to calculate dose of exposure. *International Journal of Epidemiology* 1999; **28**: 353–359.
 33. **Report of an ad-hoc Committee.** Outbreak of legionellosis in a community. *Lancet* 1986; **328**: 380–383.
 34. **Harrison TG, et al.** Comparison of clinical and environmental isolates of *Legionella pneumophila* obtained in the UK over 19 years. *Clinical Microbiology and Infection* 2007; **13**: 78–85.